The relevance of theory in psychoacoustics

William M. Hartmann
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan, 48824

This essay is based on a talk at the 157th meeting of the Acoustical Society of America. It makes general remarks about the role of theory in science with particular reference to hearing science. Criteria for evaluating theories are suggested in terms of correctness, elegance, and utility.

I. INTRODUCTION

In ordinary conversation, the word "theory" normally means something that is contrary to "facts," or, if not contrary to facts, something that is indifferent to facts. It is said, "That idea is fine in theory but it fails in practice."

An example of that popular view is a quote, attributed to Yogi Berra, "In theory there is no difference between theory and practice, but in practice there is. "

In the scientific sense, theory has a more central and more practical role. One of its roles, though not the only role, is actually to be successful in practice. In science, theories are expected to work.

Theory is the consolidation of experimental facts into a causative, conceptual framework that serves as a guide to further experiments. There are three elements to that sentence:

1. Consolidation of experimental facts.

2. Causative framework, i.e. a mechanism or process held to be responsible for the facts ... a statement about how things work.

3. Predictions for future experiments based on the mechanism.

In connection with item 1, consolidating experimental facts, Licklider wrote, "In the field of hearing ... there are so many experimental facts that systematic organization of them without the aid of models amounts to cataloging."

It is amusing to think that Licklider wrote those words about 50 years ago. Since then there have been 600 issues of JASA, all full of new experimental facts in the field of hearing. The need for models or theories is even greater now than it was then. (The reader will note that this essay does not distinguish between the words "theory" and "model," though in other contexts one could say that a model is a specific instance of a theory.)

A scientific theory may have its origin in observation, but it is more than observation; it also postulates a mechanism that makes assumptions. Fans of Monty Python will remember the famous theory of the brontosaurus by Ann Elk. Her theory was, and I quote, "All brontosauruses are thin at one end, much, much thicker in the middle, and then thin again at the other end."

This statement about the brontosaurus fails to be a theory because it is only an observation, and not a very helpful one at that. It fails to involve mechanism and it makes no predictions beyond the scope of the original observation.

Theories may be well founded or weak, they may be elegant or ugly, but they are essential prologues to experiment because theories establish the expectations that motivate the choice of experiments to be done and the choice of parameters to be used in those experiments.

Of course, there are other opinions concerning the role of theory. Sherlock Holmes, said, "It is a capital mistake to theorise before one has data. Insensibly one begins to twist fact to suit theories instead of theories to suit the facts."

I think that most scientists would disagree with the celebrated forensic wizard of Baker Street. Collecting data efficiently, indeed, the matter of *which* data to collect in the first place depends on theorizing before one has data. The twisting of facts to suit the theories is an occupational hazard for us, but our scientific training has taught us be wary of it. Besides, in psychoacoustics, our data are often not very reliable. Maybe twisting facts is a productive thing to do until there is a complete breakdown of agreement between theory and experiment.

A dramatic testimony to the power of theory to guide experiment is seen today in the Large Hadron Collider (LHC) at CERN. It is not an exaggeration to say that this machine is a six billion dollar investment in the theory called the Standard Model, particularly the role in that theory of a never-before-seen particle namely the Higgs boson. It isn't even a particularly good theory, and much of the enthusiasm for the LHC comes from the hope that physics beyond that theory will appear. Nevertheless, the standard model is the standard for comparison, and the LHC would be unthinkable without it.

The requirement that a theory make predictions for future experiments is only a modest addition to the requirement that the theory must express a mechanism. Any non-trivial mechanism, any statement that this is the way a particular part of the world works, will make predictions. Predictions follow logically from the mechanism, though it may sometimes require considerable insight into the mechanism to expose all of the predictions.

II. QUANTITATIVE AND NOT QUANTITATIVE

Scientific theories may be quantitative or not quantitative. The difference depends on the nature of the data to be explained or predicted.

As an example of a not-quantitative theory, one can imagine an audiologist who measures hearing thresholds for a group of 60-year-old men. The audiologist has the theory that these individuals will show a hearing loss, and the hypothetical mechanism is the hair cell loss that occurs with advancing age. That theory is not quantitative in that it does not predict the extent or the amount of loss.

Next, the audiologist makes the threshold measurements and discovers that 90 percent of the men do have a hearing loss, but 10 percent don't. The audiologist then comes up with a revised theory that says that in a group of 60-year-old men, 90 percent will show a hearing loss. The theory is now quantitative. However, it actually fails to be a theory because there is no mechanism for the 90-percent number. That number was entirely based on observation.

Quantitative theories tend to be mathematical models, and in our field they are normally signal processing models of some kind. They may be black box models that work in the analog domain. They may be neural models that work on neural spike trains or on probabilistic representations of those spike trains.

The special role of mathematical models is that one can get more out of them than one puts in. Getting more out than was put in is the serendipitous result that emerges when the ordinary manipulation of mathematical expressions suddenly leads to insight that was never initially incorporated. Years ago Mark Klein and I were trying to relate frequency modulation detection to the frequency difference limen for pulsed tones (Hartmann and Klein, 1980). We devised a model (the correlated-differencing model) wherein the listener detects modulation by taking the difference between two internal representations of frequency measured at times separated by what the listener imagines to be half a period of the modulation. Because the listener has no clue about the modulation phase, he averages over all possibilities. The model worked well enough in that it successfully accounted for the relative detectabilities of different modulation waveforms. It also accounted for frequency modulation thresholds and the frequency difference limen for pulsed tones. The serendipitous discovery occurred when we manipulated the equations to compute the psychometric function for modulation detection and found it to be a quadratic function of the modulation index. We hadn't put that into the model. We hadn't expected that result, but it suddenly popped out of the mathematics on the page, and it agreed with the data that Jesteadt and Sims had published in 1975, and with our own data as well. One does not forget experiences like that.

The next step in a serendipitous experience like that is to ask, where in the mathematical development did the unexpected agreement arise? In this case it was in the differencing process. In this way, our own mathematical model taught us something that we didn't know before. What we knew before was how to do the algebra.

III. THEORY EVALUATION - CORRECTNESS

There are good theories and there are bad theories. Theories need to be evaluated, and there are several criteria for evaluation. The most obvious criterion is correctness – a theory needs to be self-consistent and it should make correct predictions.

There is a philosophical concept, noted by Karl Popper and recently reiterated by Steven Hawking, whereby a theory can never be proved to be true because no matter how many observations it successfully explains, it may fail to explain the next observation. According to that philosophy, a theory can only be disproved, and one exception to the theory is enough to disprove it. By that standard, the audiologist's first theory about hearing loss in sixtyish men has been disproved because 10 percent of the the observations failed to support it.

In practical terms, theories about human or animal capabilities or behavior must be judged according to a different standard than theories about hard sciences. We are likely to say that the audiologist's theory is pretty good because it agrees with 90 percent of the observations. Our standard for the correctness of a theory in this area is not absolute. The advantage of relaxing standards in this way is that it gives us the opportunity to say that a theory that agrees with observation 90 percent of the time has an advantage over a theory that works 80 percent of the time. There is no such quantitative measure if a theory is either right or wrong.

There are definitely different levels of expectation about correctness. We know Lord Rayleigh for his acoustical contributions, but his Nobel prize (1904) was won for the discovery of the chemical element, argon. Rayleigh's discovery arose because his attempt to measure the density of nitrogen gas by two different techniques led to a discrepancy of one part in 1000. As it happened, the science was well enough developed that a discrepancy of such a size was enough to lead to a minor revolution. It really was a revolution because at that time there was no place for argon, or for any of the inert gases, in the periodic table of elements. It goes without saying that those of us who work in hearing science would be delighted to find that our discrepancies were only one part in a thousand. Different standards apply.

Even in the hardest of the physical sciences there is wiggle room for theories that don't quite work all the time – in classical mechanics for example. We know that Newton's laws of motion are valid and useful for almost all of our daily engineering. Newton's laws are self-consistent; they are elegant; and they work. The celestial mechanics used to compute the trajectories of Apollo astronauts to the Moon and back was based on Newtonian mechanics. But we know that Newtonian theory is only an approximation that has been subsumed by relativistic mechanics. When the velocities involved in mechanics approach the speed of light, Newton's laws no longer work. Then it is necessary to use the laws of special relativity. Special relativity correctly describes mechanics at speeds near the speed of light, and what's more, the laws of special relativity reduce to Newton's laws for lower speeds. The advance of relativity is practical too. It was required to make the global positioning system work.

What then should one say about Newton's laws? The easiest thing to say is that Newton's laws are wrong. An exception has been found for those laws and that makes Newton's theory wrong. However, that is not what is usually said about Newton's laws. Instead, scientists normally say that Newton's laws are a good theory so long as they are applied in their correct domain.

Saying that a theory is valid so long as it is limited to its domain of applicability is dangerously close to saying merely that a theory is right when it is right. In order for such a statement to be meaningful it is necessary for the theory to precisely state what that domain of applicability is. In the case of classical mechanics, Isaac Newton did not have a clue about the speed of light, or any limiting speed, and he was in no position to qualify the applicability of his laws of motion. Neither were any other scientists for the next 150 years. Over the course of the 20th century we learned to understand the domain of applicability of classical mechanics.

The history of hearing science also shows the progressions of models for which the domain of applicability only slowly emerged. Based on his psychoacoustical experiments and anatomical observations, Helmholtz hit upon the idea of the cochlea as a tuned system, with different places in the cochlea resonating at different frequencies like piano strings. That theory is more or less right, so long as you don't expect it to describe Bekesy's traveling wave. And the traveling wave theory is also correct, so long as you don't expect it to describe the active and nonlinear cochlea.

IV. THEORY EVALUATION - ELEGANCE

Another criterion for evaluating a theory is elegance. Successful theories are elegant. The American essayist H.L. Mencken wrote, "Explanations exist: they have existed for all time for there is always an easy solution to every problem - neat, plausible, and wrong."

Mencken's emphasis is obviously on the word "wrong," but it's important not to undervalue the power of "neat," because neat means elegant, and elegant theories have the power to inspire. They inspire experiments and they inspire further theories. An example from our own field is Fletcher's model of the critical band wherein decisions are made based on excitation within frequency channels. This model remains a source of inspiration, and yet we know that there are cross-channel effects. Another influential model is the Jeffress model of binaural hearing, based on the coincidence of excitation and delayed excitation. Influential or not, we know that this part of the auditory system is filled with inhibition, and inhibition is not included in the Jeffress model. Again, there is Schouten's residue theory of pitch perception, and yet we know that there is spectral dominance which discounts the importance of the residue.

Quantitatively, elegance might be defined mathematically as a ratio: the number of different things explained or predicted divided by the number of arbitrary assumptions. That definition has got at least one thing right. An abundance of exceptions or special cases is not elegant. If parsimony itself is not critical to the correctness of a theory, it is anyhow critical to the elegance of a theory.

Elegance goes further than parsimony in arbitrary assumptions. An alternative definition of elegance might be the accuracy of the predictions divided by the number of independent parameters. In this context we imagine a theory of the distant future that is a highly accurate predictor of perception along some auditory dimension. The theory requires you to know 42 parameters. These parameters are not arbitrary because, as it happens, there are exactly 42 well-established mechanisms that contribute to the perception. Our initial reaction to this theory would likely be, "Surely there has to be something simpler than that!" Although we know that perceptual processes are complicated, a model consisting of 42 mechanisms does not seem elegant. The problem is not with the model, the problem is with us. And so what passes for an elegance reflects the limitations of our own brains. If we can extrapolate current trends, it is likely that the aids to our cognition from technology, especially visualization provided by advanced computer graphics, will make us ever more tolerant of complexity and more willing to accept complicated models as elegant.

It is worth remembering that the Ptolemaic theory of the solar system was in better agreement with observations than the heliocentric theory for many years. Only the elegance of the heliocentric theory caused scientists to pursue it and ultimately include the elliptical orbits that made it agree better with observation. The solar-system example reminds us of the frequent complaint about mature theories and the addition of ever more epicycles. Some theories accrete so many special cases that they become too inelegant to be inspiring. At some point a domain of theory-driven activity becomes so complicated that it ought to be abandoned in favor of an entirely new approach. A consequence of the essential conservatism of science, and perhaps also of human nature, that we tend to realize that a theory has reached that point later rather than sooner.

V. THEORY EVALUATION - UTILITY

A third element in the evaluation of a theory is utility or usefulness. There are theories that are plausible, theories that are wrong, and theories that are not even wrong.

Plausible theories of hearing make contact with known facts or with reasonable conjectures about the mechanics and neurophysiology of living creatures. Hence, an overshoot plus saturating-response model for the growth of excitation is plausible because of what we know about adaptation on the one hand, and what we know about cell membranes on the other. Autocorrelation and cross-correlation form plausible foundations for perceptual theories because we know that the nervous system includes delays as well as logical AND operations. Hebbian learning and neural network models are inherently plausible because we can easily imagine that our personal nervous system architecture is wired up in an analogous way.

Wrong theories or implausible models posit elements with properties that contradict known facts. For instance one can imagine a model of sound localization in which sound sources are localized in space because sound waves from different directions excite different areas of the ear drum. Such a model is inconsistent with facts about sound waves in the ear canal and about ear drum vibration.

Theories that are not even wrong are often those that are not useful because they do not make testable predictions. They may be elegant mathematical constructions, but they do not currently pertain to the real business of our science. There are models that may work well in artificial intelligence applications, but there is no a priori evidence for or against them as steps in perception. If it is shown that such models successfully solve some technological problems, that fact does not necessarily tell us anything about human or animal perceptual systems. Because they begin with the mathematics and not with an image of a biological hearing system, such models run the risk of not making testable predictions, given the current state of our science.

Within the world of physics, mathematical theories without connection to physical reality or without explicit falsifiability are traditionally given little respect. It was the theorist Wolfgang Pauli who said about a theory, "It's not right; it's not even wrong." Within the field of neuroscience we need to be more open minded then that, mainly because our intuitions about how the relevant system works are less well developed than Pauli's intuitions about quantum mechanics. But times change. In our own time, considerable effort has been devoted to string theories – mathematically complicated models that make predictions that are usually untestable. Thus, there is now a role for theories that are "not even wrong." Future developments may elevate them to the status of actually being wrong. Or, as Freeman Dyson said about string theories, they may be found to be partly right and partly wrong.

VI. THEORIES AND COMPUTER ALGORITHMS

Auditory theorizing today is not the same as it was in Licklider's day. At that time a scientist would conceive of a process, express it in the form of mathematical equations, and solve the equations by some means or another. The equations and their solutions were often idiosyncratic, but at least the theorist himself or herself understood all the ingredients. The solution to the equations might have been entirely analytic or it might have involved some computational engine. In any case the solution led to a prediction which could be compared with experiment. Today, there is standardization in digital techniques for computation, for data mining, for statistics, for graphical presentation, and for communication. Because of this standardization the hearing scientist has a wealth of computer algorithms available from auditory toolboxes on the web. The scientist can mix and match processing modules such as those from Ear Lab. He does not need to write computer code if it can be instantly downloaded from the internet. Does the recent development of community theorizing represent a positive step forward?

As with most technological developments, the answer depends on how the technology is used. Using a computer procedure developed by others obviously requires the user to respect the conventions of the original developer. One can get very strange predictions from an algorithm by inputting the frequency in Hz when the algorithm expects the frequency to be in kHz. Beyond getting the units right, the user needs to understand the detailed purpose of the algorithm. What assumptions did the developer make? What were the developer's intentions? What attitudes did the developer take to matters of resolution and precision? What does the user do if the algorithm does not agree with experiment? Does the user understand the algorithm well enough to change the code? Possibly changing to new parameters, completely within the bounds of plausibility, would make a big change in the predictions. What does the user do if the algorithm DOES agree with experiment? What has been learned apart from the fact that this bit of code gets the right answer? How sensitive was that right answer to assumptions in the algorithm? What features of the algorithm were essential to the agreement? Were all of them necessary? Would a much simpler model have worked just as well? The user of code imported from elsewhere really ought to find answers to those questions.

On the positive side, the opportunity to share computer programs so effortlessly offers real advantages for theorizing in our science. Common software enhances communication. If scientist A says that he used a Meddis hair-cell model then scientist B knows in detail what A is talking about. It becomes a form of shorthand that conveys a lot of information quickly. Communication enabled by common algorithms contributes to orderly progress in the field and to a coherent archival literature. Sharing code means that the science can advance synergistically. A model can be tested in a new context never envisioned by the original developer.

However, the theorizing is done, by idiosyncratic procedures or by community effort, the principles of correctness, elegance, and utility still apply. One can no more imagine a science without a theory than one can imagine a story without a plot. Even for those of us who are dedicated experimenters, theory is the basis of what we do.

ACKNOWLEDGEMENTS

I am grateful to Erick Gallun and Nat Durlach for organizing the theory session at the 157th meeting of the Acoustical Society of America at which this paper was presented. The work was supported by the National Institute of Deafness and Other Communication Disorders.

REFERENCES

Hartmann, W.M. and Klein, M.A. (1980) "Theory of frequency modulation detection for low modulation frequencies" J. Acoust. Soc. Am. **67**, 935-946.

Jesteadt, W. and Sims, S.L. (1975) "Decision processes in frequency discrimination," JASA **57**, 1161-1168. [copy in Jesteadt file.]

Report 167, theory4.tex, 17 August 2009